FORM 1 (cDNA sequence provided):

```
    1 ATGGACAGAG TTTATGAAAT TCCTGAGGAG CCAAATGTGG ATCCGGTTTC
   51 ATCTCTGGAG GAAGATGTCA TCCGTGGAGC CAACCCCCGA TTTACTTTTC
  101 CATTTAGCAT CCTTTTCTCC ACCTTTTTGT ACTGTGGGGA GGCTGCATCT
  151 GCTTTGTACA TGGTTAGAAT CTATCGAAAG AATAGTGAAA CTTACCGGAT
  201 GACATACACC TTTTCTTTCT TTATGTTTTC ATCCATTATG GTCCAGTTGA
  251 CCCTCATTTT TGTCCACAGA GATCTAGCCA AAGATAAACC GCTATCATTA
  301 TTTATGCATC TAATCCTCTT GGGACCTGTT ATCAGATGTT TGGAGGCCAT
  351 GATTAAGTAC CTCACACTGT GGAAGAAAGA GGAGCAGGAG GAGCCCTATG
  401 TCAGCCTCAC CCGAAAGAAG ATGCTAATAG ATGGCGAGGA GGTGCTGATA
  451 GAATGGGAGG TGGGCCACTC CATCCGGACC CTGGCTATGC ACCGCAATGC
  501 CTACAAACGT ATGTCACAGA TCCAAGCCTT CCTGGGCTCA GTGCCCCAGC
  551 TGACCTATCA GCTCTATGTG AGCCTGATCT CTGCAGAGGT TCCCCTGGGT
  601 AGAGTTGTGC TAATGGTATT TTCCCTGGTA TCTGTCACCT ATGGGGCCAC
  651 CCTTTGCAAT ATGTTGGCTA TCCAGATCAA GTACGATGAC TACAAGATTC
  701 GCCTTGGGCC ACTAGAAGTC CTCTGCATCA CCATCTGGCG GACATTGGAG
  751 ATCACTTCCC GCCTCCTGAT TCTGGTGCTC TTCTCAGCCA CTTTGAAATT
  801 GAAGGCTGTG CCCTTCCTAG TGCTCAACTT CCTGATCATC CTCTTTGAGC
  851 CCTGGATTAA GTTCTGGAGA AGTGGTGCCC AGATGCCCAA TAACATTGAG
  901 AAAAACTTCA GCCGGGTCGG CACTCTGGTG GTCCTGATTT CAGTCACCAT
  951 CCTCTATGCT GGCATCAACT TCTCTTGCTG GTCAGCTTTG CAGTTGAGGT
 1001 TGGCAGACAG AGATCTCGTC GACAAAGGGC AGAACTGGGG ACATATGGGC
 1051 CTGCACTATA GTGTGAGGTT GGTAGAGAAT GTGATCATGG TCTTGGTTTT
 1101 TAAGTTCTTT GGAGTGAAAG TGTTACTGAA TTACTGTCAT TCCTTGATTG
 1151 CCTTGCAGCT CATTATTGCT TATCTGATTT CCATTGACTT CATGCTCCTT
 1201 TTCTTCCAGT ACTTGCATCC ATTGCGCTCA CTCTTCACCC ATAATGTAGT
 1251 AGACTACCTC CATTGTGTCT GCTGTCACCA GCACCCTCGG ACCAGGGTTG
 1301 AGAACTCAGA GCCACCCTTT GAGACTGAAG CAAGGCAAAG TGTTGTCTGA
 (SEQ ID NO:1)
```

FEATURES:

```
Start Codon: 1
Stop Codon:  1348
3'UTR:       1351
```

FORM 2 (transcript sequence provided):

```
    1  ATGAACACAA GACCACAACA TTCAGAAAGA ACCTCGACAA TGGACAGAGT
   51  TTATGAAATT CCTGAGGAGC CAAATGTGGA TCCGGTTTCA TCTCTGGAGG
  101  AAGATGTCAT CCGTGGAGCC AACCCCCGAT TTACTTTTCC ATTTAGCATC
  151  CTTTTCTCCA CCTTTTTGTA CTGTGGGGAG GCTGCATCTG CTTTGTACAT
  201  GGTTAGAATC TATCGAAAGA ATAGTGAAAC TTACTGGATG ACATACACCT
  251  TTTCTTTCTT TATGTTTTCA TCCATTATGG TCCAGTTGAC CCTCATTTTT
  301  GTCCACAGAG ATCTAGCCAA AGATAAACCG CTATCATTAT TTATGCATCT
  351  AATCCTCTTG GGACCTGTTA TCAGATGTTT GGAGGCCATG ATTAAGTACC
  401  TCACACTGTG GAAGAAAGAG GAGCAGGAGG AGCCCTATGT CAGCCTCACC
  451  CGAAAGAAGA TGCTAATAGA TGGCGAGGAG GTGCTGATAG AATGGGAGGT
  501  GGGCCACTCC ATCCGGACCC TGGCTATGCA CCGCAATGCC TACAAACGTA
  551  TGTCACAGAT CCAAGCCTTC CTGGGCTCAG TGCCCCAGCT GACCTATCAG
  601  CTCTATGTGA GCCTGATCTC TGCAGAGGTT CCCCTGGGTA GAGTTGTGCT
  651  AATGGTATTT TCCCTGGTAT CTGTCACCTA TGGGGCCACC CTTTGCAATA
  701  TGTTGGCTAT CCAGATCAAG TACGATGACT ACAAGATTCG CCTTGGGCCA
  751  CTAGAAGTCC TCTGCATCAC CATCTGGCGG ACATTGGAGA TCACTTCCCG
  801  CCTCCTGATT CTGGTGCTCT TCTCAGCCAC TTTGAAATTG AAGGCTGTGC
  851  CCTTCCTAGT GCTCAACTTC CTGATCATCC TCTTTGAGCC CTGGATTAAG
  901  TTCTGGAGAA GTGGTGCCCA GATGCCCAAT AACATTGAGA AAAACTTCAG
  951  CCGGGTCGGC ACTCTGGTGG TCCTGATTTC AGTCACCATC CTCTATGCTG
 1001  GCATCAACTT CTCTTGCTGG TCAGCTTTGC AGTTGAGGTT GGCAGACAGA
 1051  GATCTCGTCG ACAAAGGGCA GAACTGGGGA CATATGGGCC TGCACTATAG
 1101  TGTGAGGTTG GTAGAGAATG TGATCATGGT CTTGGTTTTT AAGTTCTTTG
 1151  GAGTGAAAGT GTTACTGAAT TACTGTCATT CCTTGATTGC CTTGCAGCTC
 1201  ATTATTGCTT ATCTGATTTC CATTGGCTTC ATGCTCCTTT TCTTCCAGTA
```

FIGURE 1A

```
1251  CTTGCATCCA TTGCGCTCAC TCTTCACCCA TAATGTAGTA GACTACCTCC
1301  ATTGTGTCTG CTGTCACCAG CACCCTCGGA CCAGGGTTGA GAACTCAGAG
1351  CCACCCTTTG AGACTGAAGC AAGGCAAAGT GTTGTCTGA (SEQ ID NO:4)
```

HOMOLOGOUS PROTEINS:

Top BLAST Hits:

	Score	E
gi|6502963|gb|AAF14527.1|AF155511_1 (AF155511) KX antigen [Mus ...	366	e-100
gi|10835267|ref|NP_066569.1| Kell blood group precursor (McLeod...	361	1e-98
gi|2135606|pir||I39294 McLeod syndrome-associated protein XK - ...	358	8e-98
gi|3183551|sp|P51811|XK_HUMAN MEMBRANE TRANSPORT PROTEIN XK (KX...	358	1e-97
gi|4759330|ref|NP_004668.1| Testis-specific XK-related protein ...	76	8e-13

BLAST to dbEST:

	Score	E
gi|1891549 /dataset=dbest /taxon=9606 ...	383	e-104

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

Expression information from BLAST dbEST hits:

gi|1891549 Germinal center B cells

Expression information from PCR-based tissue screening panels:

Mixed tissue

FIGURE 1B

FORM 1:

```
    1 MDRVYEIPEE PNVDPVSSLE EDVIRGANPR FTFPFSILFS TFLYCGEAAS
   51 ALYMVRIYRK NSETYRMTYT FSFFMFSSIM VQLTLIFVHR DLAKDKPLSL
  101 FMHLILLGPV IRCLEAMIKY LTLWKKEEQE EPYVSLTRKK MLIDGEEVLI
  151 EWEVGHSIRT LAMHRNAYKR MSQIQAFLGS VPQLTYQLYV SLISAEVPLG
  201 RVVLMVFSLV SVTYGATLCN MLAIQIKYDD YKIRLGPLEV LCITIWRTLE
  251 ITSRLLILVL FSATLKLKAV PFLVLNFLII LFEPWIKFWR SGAQMPNNIE
  301 KNFSRVGTLV VLISVTILYA GINFSCWSAL QLRLADRDLV DKGQNWGHMG
  351 LHYSVRLVEN VIMVLVFKFF GVKVLLNYCH SLIALQLIIA YLISIDFMLL
  401 FFQYLHPLRS LFTHNVVDYL HCVCCHQHPR TRVENSEPPF ETEARQSVV (SEQ ID NO:2)
```

FORM 2:

```
      1  MNTRPQHSER TSTMDRVYEI PEEPNVDPVS SLEEDVIRGA NPRFTFPFSI
     51  LFSTFLYCGE AASALYMVRI YRKNSETYWM TYTFSFFMFS SIMVQLTLIF
    101  VHRDLAKDKP LSLFMHLILL GPVIRCLEAM IKYLTLWKKE EQEEPYVSLT
    151  RKKMLIDGEE VLIEWEVGHS IRTLAMHRNA YKRMSQIQAF LGSVPQLTYQ
    201  LYVSLISAEV PLGRVVLMVF SLVSVTYGAT LCNMLAIQIK YDDYKIRLGP
    251  LEVLCITIWR TLEITSRLLI LVLFSATLKL KAVPFLVLNF LIILFEPWIK
    301  FWRSGAQMPN NIEKNFSRVG TLVVLISVTI LYAGINFSCW SALQLRLADR
    351  DLVDKGQNWG HMGLHYSVRL VENVIMVLVF KFFGVKVLLN YCHSLIALQL
    401  IIAYLISIGF MLLFFQYLHP LRSLFTHNVV DYLHCVCCHQ HPRTRVENSE
    451  PPFETEARQS VV (SEQ ID NO:5)
```

FEATURES:
Functional domains and key regions:

```
[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

Number of matches: 2
      1    302-305 NFSR (SEQ ID NO: 7)
      2    323-326 NFSC (SEQ ID NO: 8)

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
cAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2
      1      59-62 RKNS (SEQ ID NO: 9)
      2    169-172 KRMS (SEQ ID NO: 10)

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 6
      1      64-66 TYR
      2    137-139 TRK
      3    157-159 SIR
      4    252-254 TSR
      5    264-266 TLK
      6    354-356 SVR

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 3
      1      17-20 SSLE (SEQ ID NO: 11)
      2      18-21 SLEE (SEQ ID NO: 12)
      3    431-434 TRVE (SEQ ID NO: 13)
```

FIGURE 2A

[5] PDOC00007 PS00007 TYR_PHOSPHO_SITE
Tyrosine kinase phosphorylation site

126-133 KEEQEEPY (SEQ ID NO: 14)

[6] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 4
1 215-220 GATLCN (SEQ ID NO: 15)
2 321-326 GINFSC (SEQ ID NO: 16)
3 343-348 GQNWGH (SEQ ID NO: 17)
4 350-355 GLHYSV (SEQ ID NO: 18)

[7] PDOC00029 PS00029 LEUCINE_ZIPPER
Leucine zipper pattern

100-121 LFMHLILLGPVIRCLEAMIKYL (SEQ ID NO: 19)

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	36	56	1.443	Certain
2	74	94	2.084	Certain
3	102	122	0.920	Putative
4	181	201	0.811	Putative
5	208	228	1.744	Certain
6	273	293	1.234	Certain
7	312	332	1.785	Certain
8	366	386	0.828	Putative
9	389	409	1.497	Certain

FIGURE 2B

BLAST Alignment to Top Hit:

```
>gi|6502963|gb|AAF14527.1|AF155511_1 (AF155511) KX antigen [Mus
          musculus]          Length = 446
 Score =  366 bits (930), Expect = e-106
 Identities = 179/411 (43%), Positives = 265/411 (63%), Gaps = 11/411 (2%)

Query: 33  FPFSILFSTFLYCGEAASALYMVRIYRKNSETYRMTYTFSFFMFSSIMVQLTLIFVHRDL 92
           FP S++ S FL+  E A+ALY+   YR   +      T  F +    +VQ TL+FVHRDL
Sbjct: 3   FPASVIASVFLFVAETAAALYLSSTYRSAGDRMWQVLTLLFSLMPCALVQFTLLFVHRDL 62

Query: 93  AKDKPLSLFMHLILLGPVIRCLEAMIKYLTLWKKEEQEEPYVSLTRKKMLI-DGEEVLIE 151
           ++D+PL+L MHL+ LGP+ RC E    Y    + ++ EEPYVS+T+K+ +  DG    +E
Sbjct: 63  SRDRPLALLMHLLQLGPLYRCCEVFCIYC---QSDQNEEPYVSITKKRQMPKDGLSEEVE 119

Query: 152 WEVGHSIRTLAMHRNAYKRMSQIQAFLGSVPQLTYQLYVSLISAEVPLGRVVLMVFSLVS 211
            EVG +   L  HR+A+ R S IQAFLGS PQLT QLY++++   +  GR  +M  SL+S
Sbjct: 120 KEVGQAEGKLITHRSAFSRASVIQAFLGSAPQLTLQLYITVLEQNITTGRCFIMTLSLLS 179

Query: 212 VTYGATLCNMLAIQIKYDDYKIRLGPLEVLCITIWRTLEITSRLLILVLFSATLKLKAVP 271
           + YGA  CN+LAI+IKYD+Y++++ PL  +CI +WR+ EI +R+++LVLF++ LK+  V
Sbjct: 180 IVYGALRCNILAIKIKYDEYEVKVKPLAYVCIFLWRSFEIATRVIVLVLFTSVLKIWVVA 239

Query: 272 FLVLNFLIILFEPWIKFWRSGAQMPNNIEKNFSRVGTLVVLISVTILYAGINFSCWSALQ 331
            +++NF      PWI FW SG+  P NIEK  SRVGT +VL  +T+LYAGIN  CWSA+Q
Sbjct: 240 VILVNFFSFFLYPWIVFWCSGSPFPENIEKALSRVGTTIVLCFLTLLYAGINMFCWSAVQ 299

Query: 332 LRLADRDLVDKGQNWGHMGLHYSVRLVENVIMVLVFKFFGVKVLLNYCHSLIALQLIIAY 391
           L++ + +L+ K QNW  + ++Y  R +EN +++L++ FF   + +  C  L+ LQL+I Y
Sbjct: 300 LKIDNPELISKSQNWYRLLIYYMTRFIENSVLLLLWYFFKTDIYMYVCAPLLILQLLIGY 359

Query: 392 LISIDFMLLFFQYLHPLRSLFTHNVVD----YLHCVCCHQHPRTRVENSEP 438
              I FML+F+Q+ HP + LF+ +V +      L C C       R ++SEP
Sbjct: 360 CTGILFMLVFYQFFHPCKKLFSSSVSESFRALLRCACWSS---LRRKSSEP 407
(SEQ ID NO:6)
```

ALIGNMENT OF FORM 1 AND FORM 2:

```
>FORM 2
          Length = 462 (Length of FORM 1 = 449)
 Score =  900 bits (2301), Expect = 0.0
 Identities = 447/449 (99%), Positives = 447/449 (99%)

FORM 1: 1   MDRVYEIPEEPNVDPVSSLEEDVIRGANPRFTFPFSILFSTFLYCGEAASALYMVRIYRK 60
           MDRVYEIPEEPNVDPVSSLEEDVIRGANPRFTFPFSILFSTFLYCGEAASALYMVRIYRK
FORM 2: 14  MDRVYEIPEEPNVDPVSSLEEDVIRGANPRFTFPFSILFSTFLYCGEAASALYMVRIYRK 73

FORM 1: 61  NSETYRMTYTFSFFMFSSIMVQLTLIFVHRDLAKDKPLSLFMHLILLGPVIRCLEAMIKY 120
           NSETY MTYTFSFFMFSSIMVQLTLIFVHRDLAKDKPLSLFMHLILLGPVIRCLEAMIKY
FORM 2: 74  NSETYWMTYTFSFFMFSSIMVQLTLIFVHRDLAKDKPLSLFMHLILLGPVIRCLEAMIKY 133

FORM 1: 121 LTLWKKEEQEEPYVSLTRKKMLIDGEEVLIEWEVGHSIRTLAMHRNAYKRMSQIQAFLGS 180
           LTLWKKEEQEEPYVSLTRKKMLIDGEEVLIEWEVGHSIRTLAMHRNAYKRMSQIQAFLGS
FORM 2: 134 LTLWKKEEQEEPYVSLTRKKMLIDGEEVLIEWEVGHSIRTLAMHRNAYKRMSQIQAFLGS 193

FORM 1: 181 VPQLTYQLYVSLISAEVPLGRVVLMVFSLVSVTYGATLCNMLAIQIKYDDYKIRLGPLEV 240
           VPQLTYQLYVSLISAEVPLGRVVLMVFSLVSVTYGATLCNMLAIQIKYDDYKIRLGPLEV
FORM 2: 194 VPQLTYQLYVSLISAEVPLGRVVLMVFSLVSVTYGATLCNMLAIQIKYDDYKIRLGPLEV 253

FORM 1: 241 LCITIWRTLEITSRLLILVLFSATLKLKAVPFLVLNFLIILFEPWIKFWRSGAQMPNNIE 300
           LCITIWRTLEITSRLLILVLFSATLKLKAVPFLVLNFLIILFEPWIKFWRSGAQMPNNIE
FORM 2: 254 LCITIWRTLEITSRLLILVLFSATLKLKAVPFLVLNFLIILFEPWIKFWRSGAQMPNNIE 313
```

FIGURE 2C

```
FORM 1: 301 KNFSRVGTLVVLISVTILYAGINFSCWSALQLRLADRDLVDKGQNWGHMGLHYSVRLVEN 360
        KNFSRVGTLVVLISVTILYAGINFSCWSALQLRLADRDLVDKGQNWGHMGLHYSVRLVEN
FORM 2: 314 KNFSRVGTLVVLISVTILYAGINFSCWSALQLRLADRDLVDKGQNWGHMGLHYSVRLVEN 373

FORM 1: 361 VIMVLVFKFFGVKVLLNYCHSLIALQLIIAYLISIDFMLLFFQYLHPLRSLFTHNVVDYL 420
        VIMVLVFKFFGVKVLLNYCHSLIALQLIIAYLISI FMLLFFQYLHPLRSLFTHNVVDYL
FORM 2: 374 VIMVLVFKFFGVKVLLNYCHSLIALQLIIAYLISIGFMLLFFQYLHPLRSLFTHNVVDYL 433

FORM 1: 421 HCVCCHQHPRTRVENSEPPFETEARQSVV 449 (SEQ ID NO: 2)
        HCVCCHQHPRTRVENSEPPFETEARQSVV
FORM 2: 434 HCVCCHQHPRTRVENSEPPFETEARQSVV 462 (Position 14 to 462 of SEQ ID NO:5)
```

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
CE00306	E00306 Membrane_transport_protein_XK	390.8	1.3e-113	1

Parsed for domains:

Model	Domain	seq-f	seq-t		hmm-f	hmm-t		score	E-value
CE00306	1/1	31	416	..	1	384	[.	390.8	1.3e-113

FIGURE 2D

```
   1 TATTATTATT ATTATTAAGA CGTAATCTTG CTCTGTTGCC CAGGCTGGAG
  51 TGCAGTGGCG TGATCTCAGC TCACTGCAAC CTCTGCCGTC CGGGTTCAAG
 101 TTTTTCTCCT GCCTCAGCCT CCTGAGTAGC TGGGATTACA GTCACGCACC
 151 ACCACGACCA GCTGATTTTT GTATTTTTAG TAGAGATGGG GTTTCACCAC
 201 GTTGGCCAGG CTGGTTTCGA ACTCCTGACC TCAAGTGATC TGCCTGCCTC
 251 AGCCTCCCAA AGTGCTGGGA TTACAGGCGT GAACCACTGT GCCTGGCCTT
 301 CATCTATATT ATTACCAGGA GGCAGATGTG TTCTCTTTTT CTCTGAGGTT
 351 TAGAATTATG CAAATGAAGA TATGAAAACA AAAGCTCAGT GAGGTGGGGA
 401 GGATTACACT TAAGAATACA GGTAATTTTC AAAGCTCTTT AAGACACCCC
 451 TCTCAGTTTT TACTAACAGC TCTCTCTTGG CTCTTTGCCA GTCTGTTTAG
 501 AATTTGGCAC CTCTTCATAA CCTTTCAACC AAAGACCTGT AAGTTCATTC
 551 TAAAGCTCCT ATCCTGGCCT CATTTTGCAA GTGGAGAAAT CAAGGCATAA
 601 AATATGAGCT TTCAGTGTCT GTGGGCTGAC CTTGAGTCTT GACCTTTATC
 651 CTGTTCTATC TTCCCTCCGC CGAAAACTCT GACCCTATTC CTCCCAGGTT
 701 CCCCCTTCAT GATATTATCT GGAGGGCAAT AGGACCTAGG GAGGTTCCAC
 751 CCTGCGGCGG AGGGAGACAC ACCTGCCTAA CAGCGTGGGT AGAGTGAGTG
 801 TTGAAGCAAG TCACTTAACT AGTTAGGGAG GGCGGGGTAG AAGTGGGGGC
 851 CTGCTGCTCC TAGGGAGGAG TAAAGCTGTG GCTCCTGCCT GGGTCTGGAG
 901 GTGGTGGTCA GAAGTGCTTC TGAAGAGCGG CCCAAGCCCC TTTTTGTCCC
 951 GCCACTCCAC AACGAGCATC CCTCGGCTGG CCGCCTGCCC GGGAACTCTC
1001 CGGCTGGTTT TGTTTGGCCG CAGCCGTCCC GCCCATCTCG CCCGCCCCCG
1051 CCGTCCCGGT GCCTTAGTTT TTGAAGCTGC CGACCTCTCG CAGCTGGAAT
1101 CGCAGACCAG GCAGGACCCT GGCAGCAGAC GGCGTCCAAG AGTTTGGCGA
1151 CCTCCGTCCA GCCAGGTTGG CGCCCCGCAC ATCGTGCCTC TCACTAGCAA
1201 AGTTTCTCCG AGGAGAAGCA GCCCCTCCAG CCTTTTCTTC ATCCTGTAGA
1251 GCGAGCGCGC TCTGCTTCTG TCCCTCAACA CTGCATTCGG AGACAGGGTG
1301 GTGACAATAC TCCACTCCCG GGCCAGGCGG TCTTGGGGGC GGGGCTTGGG
1351 GGAATCCGAG GAGCTATCCT GAGAACCCTG GACTCGGCAA AGGTCCTGAG
1401 AGCGCGCAGG TGAGCGGGCC AGCTGATAGC TACAGCCTAG CAATAGCTAG
1451 GATACCTAGG CACTGAACTG AATCCCCTCT TCTGCCCTCC TTCTTCTGCG
1501 CCCGCTCTTC TGCCCTGGCT CAGCTCTCCG CTGACTTGAG AGGACACACT
1551 GGTCAGGACT CTTTGTGAGG AGCTGCTGAG TGTCGGTGCC CCCGACAGAT
1601 CGGCTACACC CTGCCTGAGG GGCTGCGAAA GGAGCCGCCA CGGAAGCCGC
1651 TGTTCTCATG ACTCTTCACG TCCCTGGAGT TGGACTCTGG ATGGGGCGCT
1701 GGGATGCTTG CTTTTGTCTT GTTCAAGTTT CACAGCAAGT ATGTTGACGA
1751 TTGGAATCGG GGCCAATCAA GAGTCAAGTT CAAAGTGGTA CTCCTGGGCT
1801 TTCCATCCCA GACTCCAAGT CGAATCTGAG TCTAGAAGAG AGCGGTTTCT
1851 TGCTCTAACT AGTGAATCTC TGTTCCCAAA CTGGACTTGA CAGAGCTCTC
1901 CTCACCTATA CTTGGACTGT AGCGGCCATA GGGTTCTCTT GGGGATGGGT
1951 GGGAGGGTGC TATGAACACA AGACCACAAC ATTCAGAAAG AACCTCGACA
2001 ATGGACAGAG TTTATGAAAT TCCTGAGGAG CCAAATGTGG ATCCGGTTTC
2051 ATCTCTGGAG GAAGATGTCA TCCGTGGAGC CAACCCCCGA TTTACTTTTC
2101 CATTTAGCAT CCTTTTCTCC ACCTTTTTGT ACTGTGGGGA GGCTGCATCT
2151 GCTTTGTACA TGGTTAGAAT CTATCGAAAG AATAGTGAAA CTTACTGGAT
2201 GACATACACC TTTTCTTTCT TTATGTTTTC ATCCATTATG GTCCAGTTGA
2251 CCCTCATTTT TGTCCACAGA GATCTAGCCA AAGATAAACC GCTATCATTA
2301 TTTATGCATC TAATCCTCTT GGGACCTGTT ATCAGGTGAG CAACTTTTAA
2351 ATCTTTTCCT TACCCCCCTA ACCCCACCCC AGACTTGGGC AGAGAAAGAT
2401 GAAAGATTTA CAAGATGGAT ACTATGGCTC TAATCAATTC TCTCATTTCC
2451 TCCCACTCTC GGCTTCCCTG TCTACCATTC AGAAAACTTA CCTGAAATCT
2501 TAAATGCCAC CATGATGAAC ATGTGGTATG TACTTGTGTT CCAAAACAAT
2551 GAACGATGCT ATTTGGGCTG TGTAAACTAG AATGGGAACA ACAAGACGTG
2601 ATCACCCTGT GCATGAAGGC CATAGCTGCA GAGTGTGTAA TTTTATTTAA
2651 AAAAATTTTT TTTTCTGAGA CAAGGTCTTG CTCTGCCTCC CAGGCTACAG
2701 TGCAGTGGTG CGATCATGGC TCACTGCAGC CTTGATCTCC TGGGATCAAG
2751 CGAACCTCCC ACCTCAGCCT CCAAGTAGCT GGGACCAAAG GAATGTGTCA
2801 CCATGCCTGG TTAATTAAAA AAAAATTTTT ATAGGCCGGG TGTGGTGGCT
2851 CATGCCTGTA ATCCCAGCAC TTTGGGAGGC TGAGGCGGGT GGATCACCTG
2901 AGGTCAGGAG TTCAAGACCA GCTGGCCAAC ATGGTGAAAC CCCTGTCTCT
2951 ACTAAAAATC AGCTGGGTGT GGTGGCGCAT ATCTGTAATC CCAGCTACTC
3001 TGGTGGCTGA GGCAGGAGAA TCACTTGAAC CCGGAAGGTA GAGGTTGCAG
```

FIGURE 3A

```
3051 TGAGCCAAGA TCGGTGCCAC TGCACTCCAG CCTGGGCGAT AGAGTGAGAC
3101 TCCATCTCAA AAAAAAAAAA ATTTTTTTTG TAGAGACGGG ATCTCGTTAT
3151 GTAGACTGGG CTCAAGTGAT CTTCCTGCCT CAGCCTCCCA AAGTGAGCCA
3201 CCACGCCTGG TCTGAGTGTG TAATTTTGAC TCTACCTTTT TGGATGCTTT
3251 GTAAATTGGA TAAAAGTTTC TTTACCCTGA GCTGCTTGGG CTGGTGCTAC
3301 TGCCATTTTC AAATTTTCCA GAGTAATGTG ACATCTGGAA ACTATTTTAA
3351 ACCATCTGTG GTAATCTGTA CCCCAACCCA ATATAGTTCA GTTCTCTGTC
3401 GGTTTATCAG TTTCCTATTT ATCTCTTTGT ATATTTCTGC AATAAAGATA
3451 CGAAGTTGGG AGGGGGCAAA GGAAGGCAGT TCATCTCTCT ATGTGGATGC
3501 AGTAGCACAA TTTAATAGTA TCAAGTATTT CCATTCAGAT TGCCTTGAAG
3551 TGGAAAGAAT GCACTTAATC CTAGCGAGAT AGGCACCTGT GTCAACAGTC
3601 TCATCTGGAT GCTATGGGGT TTTCAAGGTA GAGAGATGTT GCAAAACTTA
3651 TGAGTTCAGG AGTAAGGAAT GGACCAAGTT TGTCTTGATT GCGAGAGAGG
3701 CAGACAACTG CAGTCAGCCG AGGAATATGG GTCAGAGTGT TGCAATGGGA
3751 AGATACCTCA TCATTAGACA ACTAAAAAGT CTGTGAAACT AATTAAGGAT
3801 GGAACTCACT CCTTTATAAA ATTTCATATC TGTACACATG TATAATTTTT
3851 ATTTGTCACT TATACCTCAA TAAGGCCAAA AAAATTTTTT ATCAATAAAT
3901 TTTTAAGTGG GGAGGAATCG ATTAGGCTCT ATCAGAGAGA ATATGGGATA
3951 TCAATGGAAA CAGTGGCCTG AAATTTGGAG TCTAGTCTTC CGCCTGTCAT
4001 TGACTGGTTG TGTGTTCTTG GTAAAATCTC TGAAGATGGC TTCACAGGAA
4051 GGCATATAGA GTTCCCTCAT CTGTAAAGCA AATGGGTTAG TCTAAATCAT
4101 GGGTCTCAAA CTCAAACACT TGCAGGGACC AGGCAGGTAT CATAAATGAA
4151 TGAAGCAGGC CTAGTATAAG AAAAAACAGT AGCCTTGTGT GAGATGATAA
4201 ATGGAAACAA AGTCTCAGAG AAATACTGAG GAGTAGTGAG TACCATGGTA
4251 ATCTGAAATC TTCATGACCT GCCTGAAGGA GGTAGCCCCT CTAGAGCCCT
4301 GGCGCATTGT TTCCATGTTG GAATTCAGAC CCAGTATTGC CAGATCCACT
4351 AACTTTTCGG GAGATGCTCC CAAGACAGGA TTTTTATATG AAATGTCATG
4401 ATTTTAAATT TTCACAGCTG ACTAAAACAA TAACAACAAC AACACAGGAT
4451 GGACCAAACC ATATCTGTTG GTCAGATATA ACTCAGCTGG CCTATATGCA
4501 TCTTTGGACT GGGTGATGTA AAGGTCCTTT ACGGTTCTAA ATCTTTGAAG
4551 TTAAGCTGTA AAAGGAAGAC CTCATCTTGA CCTTGAAACC AAGAAATTTA
4601 AAGTTGTGAC TACAGGAGCA AATAAACCAT TCATCCCTCC TTTTTCAAAT
4651 ACAATATATT GAGTTAACCA ATCGAAAACT CTCAAGATAC AAATTTCAGA
4701 AAGTACCCAG CTGCACCCTC CCCTCTTTTT GACTTCCTTT GTTTGCTTTG
4751 TGAACCCTCT GTGTAGAGTG TTGAGTACTG TTTTTCATTT TTGTTGTTTA
4801 GCTTCCACTA GAAATGATTG GGAAGCATTT ATAACCTCAG GCAGCTTAGC
4851 CCACAGCAGA GAAAAGATAA AAACTCATAA ATTATACTCT GGATTCGCTT
4901 ATTTTCAAGG CCAATTACTT GTTAGATAGG TAGGAACTTG ATTAGTGTTA
4951 TCAGGCACAT GAAGGTGCTT GTAGAGTCTG GGTGCCTTAC ATGAAATGCA
5001 AGCATACTTC CGAAATGAAA ATGTACTCTA ATTTATTGAA GCTTATAAAT
5051 GGACAAACAC CCTTACTTAA ACCAGAAAAT AGCCCTGAGA ATAGAAACAG
5101 AACATTTATG TAAATGTAAA CGGAACATTT CATGCCACCA CCTTCTCCAA
5151 TACTGTTCTC CAATTTAGCA ATAGTACTGA TGGGTTGGGG TTAAAATCTA
5201 AAATTTTTCA TTGAAAATGC ACTTATGCAG AACAAGAATA GGAAAAAAGT
5251 GTTGCTTTTT CTTCTCTGTT CTTTCTTTGC ATCTTTTTCT TTCCCAGGTC
5301 TTAGAGTTTG TCCCTAGAAG GTGACAATTT CAAACTACAT GCTTCAGAGT
5351 GGTACACATG CATCAGTCTT AGGGTGATCT ATGGAGACTG GCAGCCAGCA
5401 TATGTTCCAA ATTTTCCTAT CAGGAACTAA AGGCTAGAGA GCATATCAAC
5451 CTCTGGGCTT GTCTTTGGTC TACTTTTCTG TTAAATTTCA TTGCTGTTAT
5501 TATTATCCTC TCCTCCCATA ATTGCTTACC CTGTATTATT TTCTTCCTTC
5551 TTATTCTTTC ATTTACTCAG CAAATATTTC TCAAATACCT ACTAAGTGAT
5601 AAGAGCTGTA AACAAGATAA ATACAACCCT TGACCTCAGT CTCTTGGGCA
5651 AGACGTGTTA ATGTCCACTA CAAATGTTCT TACTAGTCAT AAGTAGTCCA
5701 CAGTTTTTAT TCATTAAAGG TGAGTGGCGA AGTGGTAACT CAGGTGTTCC
5751 AGTAACAAGA ATGTTCTAGT TGCTTCTCTT CCACTTACCA CATCAGAACT
5801 GCTAAAGACT TCTGATTTGT ATGGGGGAGG TGGGAGGGGC AGAGCAGGAA
5851 ATGTCATCTT ACCCTTATTC CAAGGATGAT AGGCTTTCAT AAGGATGTTT
5901 TTCTCTTCGT AAAGAAAGAA TCCAGTTTAA AAGGCTTTTG TCCACAAACA
5951 GGACAAGAGG CACAAAAAGT AACTATTACA GTGATCTTTC GAGGGCCTAG
6001 TTATGTAGTT CATTCAGGTT TGAGTTGTCG TCTTTTAAGT ACTTTTGTTG
6051 CTTTGATGGC TTCCTGTGTA TATGAGATAT TTTTTTTCCT CTGATCTGTC
```

FIGURE 3B

```
6101 CCAAGACTTT TTGGCTGAGA TATGGTTGTG AGCCCTTTCT TGAAAAAGCA
6151 GAATCTGGCC AGGCGCAGTG GCTCATGCCT GTAATCTCAG CACTTTGGGA
6201 AGCTGAGGTG GGTGGATCAC CTGAGGTCAG GAGTTCAAGA CCAGCCTGGC
6251 CAACATGGTG AAAACCCGTC TCTACTAAAA ATACAAAAAA AAAAAAAACC
6301 TTAGCCGGAC ATGGTGGCAC ATGCCTGTAA TCCCAGCTAC TCAGGAGGCT
6351 GAGGCAGGAG AATCGCTTGA ACCCAGGAGG CAGAGGTTAC AGTGAGCTGA
6401 GATCGCGCCA GTGCACTCCA GCCTGGGCGA CAGAGCAAGA CTCTGTCTCA
6451 AAAAAAAAAA AAAAAAGAAA GAAAGAAAAA GAAAAAGCAG AATCTAAAAC
6501 TTTGGTTATG GAGCTGAATG CTTTGAGGGA GGAATGCTTT ACCTCACGAA
6551 TTTGAGGTAA GAAAACAGGG CCTTTGGAAC CTTCATTATT TTGCTAGGAA
6601 AACAGTATCG ACTTAATACC TTTGTGTTCA AGGCACTTTT CTACCTGCCA
6651 CAGGCCTATT CTTAAAAAGA CAAAACAATT CCTCGAGTCC TCAAACAAGT
6701 ACTTCTGAAA CAGTGTTCTT AGGTCAGTCG ATGACTGAAC AAAAATGGAT
6751 TTAGATTCAT GTAACTTGTA GAAGGCATGA TCCACCCTTT GACTTATGAG
6801 AAATGATCAG AACAGAAGAG AGAAAAAGAC AAAAAGTAGT GCAGGCTGGC
6851 CATGGTGTCT CACACGTGTG ATCCCAGCAC TTTAGGATCC CAGCACTTTG
6901 GGTCAAGGCA GTAGGATTGC TTGAGCCCAG GAGTTTGAGA CCAGTCTGGG
6951 CAACATGTCT AGATCTCCTC TCTACACAAA TTAAAAATAG CTGGCATGGT
7001 GGCATGCGCC TGTAGTCCTA GCTACTCAGA AGGCTGAGGT GGGAGGATCA
7051 TTTGAGCCTA GGAGGTCAAA GCTGCAATGA ATTATGATTG TGCCACTGCA
7101 CTCCAGCCAG GGTGATGGAG TAAGACCTTG TCTCAAAAAT AAAATAAAGT
7151 AGCACAACCT CCCCAAGTTA TTTTTTTCCC TCACTACAAC CTCCCTTCCC
7201 AGGACAGCTT AGTTAAGTTT GCATGATGCT TTACTTCTGC AGATGTTTGG
7251 AGGCCATGAT TAAGTACCTC ACACTGTGGA AGAAAGAGGA GCAGGAGGAG
7301 CCCTATGTCA GCCTCACCCG AAAGAAGATG CTAATAGATG GCGAGGAGGT
7351 GCTGATAGAA TGGGAGGTGG GCCACTCCAT CCGGACCCTG GCTATGCACC
7401 GCAATGCCTA CAAACGTATG TCACAGATCC AAGCCTTCCT GGGCTCAGTG
7451 CCCCAGCTGA CCTATCAGCT CTATGTGAGC CTGATCTCTG CAGAGGTTCC
7501 CCTGGGTAGA GGTGAGTGGG GTCAGGAGAG GGGAGGGCTC CAGTTAAATC
7551 AAGGGTCTTA GAAGTCTAGA CCCAAGCTGT CTAATAAACT GGCCACTAGC
7601 TTCATGTGGC TATTTAACTT AAAATTAAAT AAAATTAAAA ACTTGTTCAT
7651 TAATACTAGC TACATTTCAA GTTCTCAGCA GCCGTGTGTT GCTAGCAACT
7701 ACTGTATTGG ATGGCACAGG TATAAACATT TCCATCATCA CAGAAAGTTC
7751 TATCGGACAG CACTGGGAGA TAGTTAAATA ACTTGTGGAG TCAGACATCT
7801 CAAGCCTGCC AGATTTCTTA AACAGGTAAG CTGTTTAGAC TAAAAATGTC
7851 ACAGATAAAC CTTCTCTGGG CCCAGAAGAA GCTAGTAATA CCAGCAGTCA
7901 GTAGGATATT TTCCCTTGCC CAAAATGTTT AAATTATGCT GTTGTTTTGT
7951 TTGTTTAAGG ATGGCAGTCT TTAATAAGAG GTTCCCAAAT AGTACTGATC
8001 ATCAGAATCA TGTGATGAGC TTCTTTTTGA AATTATATTC ACTCCCCAGA
8051 CTTGAATCAA TCTTAATATG TATTTCTAAA AGGTACCCAG TTGATTTTGA
8101 TCAGCCACAT TTGGGAACCA ATGATTTAAT CATTTCTGCT AATGCCAGTG
8151 GAGAGAAAGA AAAGGAGCGT GGGCTGGGCA CGGTGGTTCA AGCCTGTAAT
8201 CCCAGCACTT TGGGAGGCCA AGGCGGGTGG ATCACAAGGT CAGGAGATTG
8251 AGACCATCCT GGCTAACATG ATGAAAACCC GTCTGTACTA AAAATACAAA
8301 AAATTAGCCG GGCGTGGTGG CAGGTGCCTG TAGTCCTAGC TACTCGGGAG
8351 GCTGAGGCAG GAGAATGGCG TGAACCTGGG AGGCGGAGCT TGCAGTGAGC
8401 CGAGATCGCG TCACTGCACT CCAGCCTGGG TGACAGAGCA AGACTCCGTC
8451 TCAAAAAAAA AAAAAAAAAA AAAAAAAAAA GCGTGGGGTT AATACTAATG
8501 AGAGTCAGGC CTGGACCAAG TTCTGACCTT CACTGTGATC TTTGGAGGAA
8551 GTTACAAAGC AATCACTGAC CTAATTTCCC ACTTGTAGAA GAGGGATCCT
8601 GAAATGAGTA AGACCTCTAG CAGAAGATGA AATGTGAGTC AGTGTTTTCA
8651 AAGTTGAGAT AAATTGTTGT TAATGAATTT TAACAGCCTG AGATTTGCTT
8701 CATCTGCTTG GGCAGGCACT GGTATAGGTG TGGGTACAGG TTTGGACCAT
8751 TTCCTATTAG ATTCTAACCC TGTTTGGCAA AGTCCCATGT CTCAAATAAG
8801 GTAAGGAGAA AATTTGCCCT CTTTTGTCTT TTTTCCCCAC TCAGAATTGT
8851 TCTTGAAGTT CTGTTGGTCT TGAAGCTTTT CACATACATA GTAGTTTGAG
8901 GAGAAAACTC TTTGGAAATG ATGATGCTTT TCCTTTAAAT CATCTAATAA
8951 AAATAGGTGT ACATTACGGC TGGGCATGAT GGCTCACGCC TGTAATCCTA
9001 ACACTTTGGG AGGCCAAGAC AGGCAGATCA CTTGAAGTTG GGAGTTCAAA
9051 ACCAGCCTGG CCAACATGTT GAAACCCCAT CTCTACTAAA AACACAAAAA
9101 AAATCAAGGA TGGGCATGGT GGCTGATACC TGTAATCCCA GCACTTTGGG
```

FIGURE 3C

```
 9151 AAGCCGAGGC AGGTGGATCA CCTGAGGTCA GGAGGTTGAG ACCAGCCTGG
 9201 CCTGGCGAAA CTCTGTCTCT ACTAAAAATA CAAAAATTAG CCGGGTATGG
 9251 TGGTGGATGC TTATAATCCC AGCTACTTGG GAGGCTGATG CATGAGAATC
 9301 ACTTGAACCT GGGAGCCGAG ATCTCACCGT TGCACTCCAG CCTGGGCAAC
 9351 AGAGCGAGAC TCTGTCTCAA AAAAAAATTC AGCCAGGCGT GGTGGTGGGT
 9401 GCCTGTAATC CCAGCTACTT GGGAGGCTGA AGCAGGAGAA TTGCTTGAAC
 9451 CTGGGAGGTG GAGGTTGCAG TGAGCTGAGA CTGCACCACT GCACCCCAGC
 9501 CTGGGCGACA GAGGGAGACT CCCGTCTCAT AAATAAATAA ATAAATAACA
 9551 AAAGTAATAC ATGCACAAAA TGACATATAA GTAATTGTAT TTGCACAGAA
 9601 AATTTCTGGA AACTATGCAA GAAACTACCT CTGCGGAGTG GGAATGAAAA
 9651 GTCAGCAGTC TTACTTTTTA AAATTCTTCT GTATGGTTTG AAAATTTTTT
 9701 TTGTGATCAT GCATTACTAG TTTTGGTCTT TATCTTTTTT TAATTACAAA
 9751 AGTCAGACAT GGTTATAGTA AAAATTAAAA ACCATACAGA ATAGATATAA
 9801 AATAGGAAAC GTAATCTCAC TCCCCAAAGA TAACCTCTGT TAATCATCCA
 9851 GTATATATCC TTCTGGACTT ATTTTTACTA TGTAAACATA AACATACATA
 9901 CAATATATAT TGTACATGTT TTTGCCCCAA AATGGACTGT ATGAAACATT
 9951 CTGTCAACAA AGTATTTTTC AAAAGTACAG TATGCCAGTA TGTCTTTTCT
10001 CAAGTTATTT ATATATACAT GTATAACAAT AATAAATATA TAATATACAT
10051 TTCCTTTATA TGAATTAGAC TATTTTTATT CTCCTAATTT TCTATTGATA
10101 GGATTCTATT GATTGTCTCA AAAAGGAAAA AAAAAGGTAG CACAACCTCC
10151 CCTAGTTATT TTTTCCCCTC ATTACAACCT CCCTTCTCAG GACAGCTTTA
10201 GTTAAGTTCC CATGATGCTT TACTTCTGCA GATGTTTGGA GGCCATGAGT
10251 AAGGACTTCA CACTGTGGAA GAAAGAGGGG CAGCAGGAGC CCTATGTCAG
10301 CCTCACAAAT TTTAATTTTT CACAAAAAAG TTGTTTCCTA ATTGCAAATT
10351 ATGCCACAGT AAACATCTTT ATAAATACCT GTGTACATGA ATGAGACTTT
10401 GTAGGATAAA TTTATAGCAG TAGAATTGCT GGGTGGAAGG ATATGTATGT
10451 TTAAAATTTT ATTGATATTG CCAAATAACT CTTCCAAAAA GATATATGAA
10501 TTTATACTCT CACCAACAGT ATACAAATGT GCCTGTTTCT GTTTCTTCAT
10551 ATCTTAAACT CAATATTCTT TATTTGTATA ATTATAAAAT AATTGGCTTT
10601 TAAAATAATT GACTTTTAAA ATAATTCGCT TTCTTTGGTT ATGAATGAAG
10651 CTGAGCATCT TTTTGTGTTT GGTCATTGTG TGTTCTGTGA ATTGCTTGTT
10701 TATATATTTT ACTCGCTTTT TCTAGTGGGT TGTCTTTTTC ATATTAATTT
10751 TTAGGAGCTA TTTACTTATT CTTGTTATTA ATCCTTTCTC TGCTGTGAAT
10801 ATGTATGCAT ATATTTGTAT AATTTTTTGC TTGTACATAC ACACATTTTA
10851 AATATGTATA TACATGTCAT ACGTGTAATA TGTGTGTGAT ATATTTAATA
10901 TCCACAATAC ACTTTGTAGT ATCTTCTGGC ATTCTGAAGT ATTACATTTT
10951 TATGTATTCA AATTCTTTAT TGCTTTTAGA TTTTGTGCCT TTCTTACAAA
11001 GGCCTATCAC ATCTCTCATC TGGTAGAACA ATTTTCCCCA ATCTTTTAAG
11051 TAGATTAATT TTCAAGATAA TTTTTTAATT CATCCTACAA AAAACAAAGC
11101 AAAATAATAA CAGCAAAAGA AAAAAACATT TCATTGAGAT TCCCATTGAG
11151 ATTTGCATCA AATTACTTAG GTTATTTTGG GAGAATTTAC ATCTTTATAG
11201 GATTGTTGGA TTTCATATTG TGAAATGATA AATCTCTCCA TTTTATTAAA
11251 TATTTTAACA TGTACCTCAG TAAAATTGTA TAGTTTTCTT CAGTAAAGTT
11301 GTATAGTTTT CTTGCTATGA GTCTTACATT TTTATAAGGC TTACTTTCAG
11351 ATGTTCTATC AGTTTTAAAA TGACCTGATT TTCTAAGTAG CAGGATAGTA
11401 TCCAGGTAAA GTAAACCCAC CTACCATACT TTTGGAAATA GGGGGATGAT
11451 GAAGATGACA AAGAATAGGA AGAAAGAGGA GGAGGGGGAG GAGGAGAAAA
11501 AGGAAAGGAA GAAGGGAAGA AGGAAGAAGA AGAAAGAAGA ACACAGCTAA
11551 AAGAATTTAC TAGGTTCTAG GCATTTCTCT AAGCCCTTTA CATGTAAATG
11601 TTTATTTAAT CTTTACCACA ACCCTATGAG ATAATTATCA TTCTCATTTT
11651 ACAGACGAAG AAACAGACGC ACTGAGAGTT TAAGTATATT CCCCCAAGGT
11701 CCCATAAGCA AAGATTGGAT TGGAATTCAG GGTGTTTGCC TCCAGAGCCT
11751 GTGTATTTTG TTCTCTTATG GCATGAGTGT ATTTGTAGGG ACACAGATTG
11801 AAAATGTTTT GACATTTATT GGAAGCATCA GGTTTTTTTC CTTCTGTTAC
11851 ACTACTAATC AATAAATGAG TTCTAATGTA AGGGAAAGCA TCGCCACACA
11901 GCTGGATGTA TGCTCTCACA TTCCCAGTTA CATAAGGTGC ATCAGCTCTT
11951 GAGGATGGGA CTGAGAATGG TTGAGAAAGA CAAGAGTCAC CACTTCAAGA
12001 GTCTCCAATT CCAGTCTCTG AGATTCCAAC ACTCTACTTA AAACTGAAGA
12051 ACTCAGAGCT GTGCTTCCTT TTGGGTTTAC ATGGGGGAAA TCTTAACTTT
12101 TCCTCCACTA AAAGTAAAAG ATTAAGTTGA CATCTCTATG GCCACCTTTT
12151 CCCTACATCA AGTGTTTTAA TAGGAACAGA AAACTCCAGC TTTCCTTTTG
```

FIGURE 3D

```
12201 GATGAGTATT CCTCAGCCAT CCCACTTCTC TTGAGAGCAC TGGATTTTTC
12251 TTAGTAATCA GCATCCTTTG ACATAAAGGA AGAAAAGGAA AGGGCCACCT
12301 GTGTCATCTA TAGTTGAGGC TTAGGTAGGT TAGGGAGCAT CCTGGCTCTT
12351 TAGGGCCACT ACTCTAACAT ATGGTTCCAT GGATGTCATG GGTGAGGCAA
12401 CAGGGTTTGG AAATTTTTGA ACTACTCTGC TGCAAACTCA GAGATTCCTA
12451 ACATATGGGG GTAATGAATT GACATTGCTG ATGACAAATA TAAGCAACTC
12501 TTGAGTATCT CAGTGAATTG AGAACTGAGG TACATAGATA TTCAGTGACT
12551 TCCAAAAGTT CCCATACAGC TGAACCAAGG ATTTCTTTCT TTCTTTCTTT
12601 CTTTCTTTCT TTCTTTCTTT CTTTCTTTCT TTCTTTCTTT CTTTCTTTCT
12651 TTCTTTTCTT TCTTTCTTTC TTTCTTTCTT TCTTTCTTTC TTTCTTCTTT
12701 CTTTCTCTTT CTTTCTCTCT TTTTCCTTTT CTTTTTCTTT CTTTTCTTTC
12751 TCTCTTTCTT TCTCTCTCTC TCTTTCTGTC TTCCTCCCTT CCTCCCTTTC
12801 TCTCTTTCTC CCTTTCTTTC TTCCCTTCCT CCTTACAGGC ATGCACCACC
12851 ATGCCCAGCT AATTTTTGTA TTTTTAGTAG AGTACCGGGT TTCACCATGT
12901 TGGTCAGGCT GGTCTTGAAC TCCTGACCTC AGGTGATCCA CCCACCTCAG
12951 CCCCGCAAAG TGCTTGGGAT TACAGGTGTG AGCCACCGTG TGCACGGCTG
13001 GAACCAAGGA TTTCTAATTA GTTTTATTTT TTATTTTTTT TCTTTTTGAG
13051 AAGGAGTCTC ACTCTGTCAC CCAGGCTGGA GTGCAGTGGG GCAATCTCAA
13101 CTCACTGCAA CCTCTGCCTC GTGGGTTCAA GTGATTCTCG TGCCTCAGCC
13151 TCCTGAGCAG CTGGGATTAC AGGCATGCCA TCATGCCTGG CTAATTTTTT
13201 TTTCTTTCTT TTTGAGACAG AGTCTCACTC TGTTGCCCAG GCTGGATCGC
13251 AGTGGTGCAA TCACGGCCCA CTATTACCTC TGCCTCCCAG GTTCAAGTAA
13301 TTCTCCTGCC TCAGCATCCC AGGTAGCTGG GAATACAGGT GCACGCCACC
13351 ACGCCTGACT AATATTTGTA TTTTTAGCGG AGATGGGGTT TCATCATGTT
13401 GGCCAGGCTG GTCTCGAACT CCTGACTTCA GGTGATCCAT CCGCCTTGGC
13451 CTCCCAATGT GCTGGGATTA CAGGCATGAG TCACCGCGCC CAGCCTAACT
13501 AGGTATTTTA TGCACCTCTC CTAATCTCAG AAGTCTTCAT TAATTCCACA
13551 AACATTTATT GAGCACCTGC TATGTTCCAG GTAATATGTT AGGCTATGGG
13601 AATACAGCAG TGAAGAAAAC ATGGTCCCTC CTGCCTTCAT GGAATTTTCA
13651 ATACACATTT TGACACATCA CTGAAGCTAA GTGTTCTAGA AACACACAAA
13701 CAATGTTAGT TCCTTGAACA AGATATACAT CAAAGAAGGG ACTTCTATTA
13751 GCAAGAGCGT TCTCTATGAG TCTCCTAAGA CTGGATTTTT TCAGATAGAG
13801 TTCTTTCCGC CTTATTCAAT GTTTGCTCCG AAGCCTGCTT CATCAGCAAA
13851 GTCTGCCTGA TACCTTTATA TGTACTCTTC TCACGTTAGT GACTTCTCAA
13901 TGTTCTAAGA CCCATGCTTT TTAAGGAAGT TTATTTTGTA TATTTATATG
13951 ATTATTAAAG TGTTACAGTA TATGTTCATC ATGAGAAATT TAGAAAATAG
14001 AGAAATGTAG AGAAAAAGAT TTCTAAAACT GATATAAGAC TATCACACAC
14051 AAAAAAAGAT ATTTTGGTTC ATTTTTTCAA TTTTTTGTGC ATCTATTTTG
14101 TTTTATTGTA TATATTCAAG GTGTACAATG TGATGTTTCG ATGTATGTAC
14151 ACATTGTGAA ATGATTACCA CAACCAAACT AATTAACACA TTCATCACCT
14201 CACATAGTTA TCATTTTTGT ACGTGTGTGT GTGTGTGTGT GTGTGTGTGT
14251 GGTAAAACTT AAGATCTACT CTCTTTAAAA ATTTCAAGTA CACAATACAT
14301 TATTGTCAAC TATAGTCATC ATGTTGTACA TTAGAGCTCT GAAACTTATT
14351 TATCTTATAA CTCTAAATTT GTAGCCTTTG ATCAAAATCC TTCTATTTCC
14401 CTAAATCCCC ATCCCCTGGT AACCACCCAT TCTACTCTGT TGCTAGGTGT
14451 TCAACTTTTT TAGATTCCAC ATATAAGTAA GACAATGCAG TATTTTTCTT
14501 TATGTGTCTA GCTCATTTCA CTTAGCATAA TGTCCTCTAG GTTCATCTGT
14551 GTTGTCACGG ATGGCAGAGT TTCTGTAATT TTATGGTTGA ATAATATTCA
14601 TACACACACA CACACACACG CACACACACA CACACACACA CAGACACACC
14651 CACCAGATTT TCTTTATCCA TTCATCTGTC AACAGATACT GAGTTTGTTT
14701 CCATATCTTG GCTATTGAGA ATAATACTAC AATGAGCATG AGAGTGCAGA
14751 TATCTCTTTG AGATACTGAT TTCCTTTAGG TATACACCCA GCAGTGGGAT
14801 TATTTGATCG TTTGGCCGTT CTGTTTGTAA TTTTTTTGGA GAACCTCCAT
14851 GCTGTTTTCC ATAATGGCTG TGTCAGTTTA TGTTCCCACA AACAGTGTAC
14901 AAGGTCCTTT TTCTTACATC CCCACCAACA CTTTTTTTTT TTAATAATAG
14951 CCATTCTAAC AGGTGTGAGG TGATATCTCA TTGTGGCTTT GATTTGCATT
15001 TTTGTGATGA TTAGTGATGT TGAACACCTT TTCATATACC TGTTGGCCGT
15051 TTGTATACCG CCTTCGGAGA AAGTCTATTC AAGTGCATGC TATTTGTTTA
15101 CATAGCTGTG ATCATATTTG CATTTGCTCT TAACTGGAGC TCTCAAGTCT
15151 CACCCGTCAT CTCTCTGGAC CTCTGGGTTA TAAGTACAGC CTTCATTACC
15201 AACATTGACT GATTGCCTGT TTTTGTTTTT GTTTTTGTTT TTAACAGTTG
```

FIGURE 3E

```
15251 TGCTAATGGT ATTTTCCCTG GTATCTGTCA CCTATGGGGC CACCCTTTGC
15301 AATATGTTGG CTATCCAGAT CAAGTACGAT GACTACAAGA TTCGCCTTGG
15351 GCCACTAGAA GTCCTCTGCA TCACCATCTG GCGGACATTG GAGATCACTT
15401 CCCGCCTCCT GATTCTGGTG CTCTTCTCAG CCACTTTGAA ATTGAAGGCT
15451 GTGCCCTTCC TAGTGCTCAA CTTCCTGATC ATCCTCTTTG AGCCCTGGAT
15501 TAAGTTCTGG AGAAGTGGTG CCCAGATGCC CAATAACATT GAGAAAAACT
15551 TCAGCCGGGT CGGCACTCTG GTGGTCCTGA TTTCAGTCAC CATCCTCTAT
15601 GCTGGCATCA ACTTCTCTTG CTGGTCAGCT TTGCAGTTGA GGTTGGCAGA
15651 CAGAGATCTC GTCGACAAAG GGCAGAACTG GGGACATATG GGCCTGCACT
15701 ATAGTGTGAG GTTGGTAGAG AATGTGATCA TGGTCTTGGT TTTTAAGTTC
15751 TTTGGAGTGA AAGTGTTACT GAATTACTGT CATTCCTTGA TTGCCTTGCA
15801 GCTCATTATT GCTTATCTGA TTTCCATTGG CTTCATGCTC CTTTTCTTCC
15851 AGTACTTGCA TCCATTGCGC TCACTCTTCA CCCATAATGT AGTAGACTAC
15901 CTCCATTGTG TCTGCTGTCA CCAGCACCCT CGGACCAGGG TTGAGAACTC
15951 AGAGCCACCC TTTGAGACTG AAGCAAGGCA AAGTGTTGTC TGATTCTATT
16001 TTCTGGGTAT TTTAGGAAGA GTTGGGAGTT GCCAAGAGTA ACCATGAAAT
16051 TGAACGAAAG GATGAGGTTC ATGGGTGAGA TACCCATCAG TACATTTTCT
16101 TGACTTTTCT GTTAAGCCTA TCAGAAGAAA GAGCAACTCC CAAATAGGTT
16151 TTATTTTCTT AAGAGTTACC ACTATGTTTG GAAACAGGGG GTATCGACTA
16201 TATAGTTGAA AGGGTCAGAA ATACCATTCA CACCCTTCTT ACCCAAGTCA
16251 ATTGGAATAA CTTGTCTTCA AACACTTTAG GCTCTCTAAA GTGACCTTCT
16301 AGCTCTGCTC ATTTGCTTGA TGCATTTCTG AGCTTTCCTG GGCTGAGCTG
16351 AAGGCCCAGA ATCCCGCTAG AATATATCCT GACTGATCAG AGGATATGAC
16401 AGCTTACCAG CTAAGAGTAC CTCCCAGGAA ACAGTCTGAC TAATGTGGAA
16451 CCTGCAACTG TCAGTGTGGC TGGGGTCTTT TTAATTCCAG TGAGAAGCTC
16501 TGGCTGAGAA GAAAATCACC ACTATTAAAA AAGCTGCTCC CCAAGCAGAT
16551 TAGCTCTCTG TTAGGATTTT ACTAGTGGCC ATTCAGCAAG GACCTCTCTT
16601 TACAGTGGCA CTTCATAGGC ACACTCTAAG GAGAAAGTGC AGAGTAGAAT
16651 TCCTTCAGGG CATAAGCCAA AATGACTCTT TTTCTCAGGG ACCTGCATGG
16701 GCCTCCAGCT TGTCTATTGG AATTGTTAAG TGAAGCCTCT CACTTAGTGC
16751 CTCATTAGCA GAGATTTCCT CCAACCCAGC TTTTCTGTGC TCTTGGTATT
16801 TTACTACTTG ATGTGGACCT CAGAGAAGCT GAACTGTAAT TGAAAATGTT
16851 TCCGATGTGT GGAAGAAATG AAGACTGCTT TGTGTCTGCT GTTGTCCTGA
16901 GTATTTCATT AATGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTATGTG
16951 TATGTGTGTA GGGAAGAAAG TAATAATGGC TGAGACATCA CCTTCATGTT
17001 GTTTGCGATT GGGATGGGTG ACTAACACTC CAAGGTAGAG TGAAGGCAGA
17051 GGAGGGAAAC AAGATCACAT TAAATCATCA TCAGTACTGG TTTCTGCCTA
17101 CAGGAGTTTA CTTTTTTTTT TTTTCCTTTT TTGAGATGGA GTCTCGCTCT
17151 GTTTTCTAGG CTGAAGTGCA GTGGTGTGAT CTTGGCTCAC TGCAGCCTCT
17201 GCCTCCTGGG TTCAAGCAGN NNNNNNNNNN NNNNNNNNNN NNNNAGTGAT
17251 CCACCCGCCT CGGTCTCCCA AAGCACTGGG ATTACAGGCA TGAGCCACCT
17301 CACGCGGCCA GGATTTTACT TTATAACAAG GAACATATGT TTATCAACCC
17351 TCTGTTCGTT CCTATACCCC CAGTGGACGA ATGCATGTCT CCTTTTCTCC
17401 TATATCTCAA TGTTTACATC TCATATCAGT TGGGTATTTT GATAGGAATG
17451 TCAGCCAGCT ACCTCTGAGG TAACCAAGGG ATTGAAGTTA CTATGGCCAC
17501 TGCCTATTGG GACCAAATAT CCCAGCATTT ACCTAACTAA TGCTTGCCCC
17551 TCACAGACCA GGAAAATTAA AAGAACTCCT AGTCGTGGCC ACCACAACAC
17601 TTCAAGAAAT TGTGAACAAT CTGACCTAGG GCTTCCTGTC CTCATCCAAT
17651 TTTACTCTTG GTAGCATGCT AAGAATTTAT CTTTAGTCAT TTCCTCTCCT
17701 CTTATCCAAT GTCAGGACAT TATGTTGAGG GAGTTCTCTC TTCTAAGTAG
17751 CAGGGCTGTT AACCAAAGTA TCTTATTTCT TGGCATGGCT AGCATGGTTT
17801 TCCCTTCATC AGCCACTGTT TGGGACTAAA AGGATTATAT ACTTAATTTG
17851 GGAGAGACTG TATGGACTTG CTTTGGAACA GTGGAGAGCT CCTTTCTTCA
17901 ACCCCAACTC CCCCATTCCA TTTTTCATGA TGAAGAGACT TAGTTATTGT
17951 CATATAAAGC TCACCTGCTG TCTTCTAACT ATGTTATTCA AGG (SEQ ID NO:3)
```

FEATURES:

Gene Structure

FORM 1:

Start: 2001

Exon: 2001-2335

FIGURE 3F

Intron: 2336-7242
Exon: 7243-7511
Intron: 7512-15247
Exon: 15248-15990
Stop: 15991

FORM 2:
Start: 1962
Exon: 1962-2335
Intron: 2336-7242
Exon: 7243-7511
Intron: 7512-15247
Exon: 15248-15993
Stop: 15991

CHROMOSOME MAP POSITION:
Chromosome 23

ALLELIC VARIANTS (SNPs):
DNA

Position	Major	Minor	Domain
2584	G	C	Intron
2655	A	T	Intron
3693	G	A	Intron
3992	G	C	Intron
6285	-	A	Intron
7066	A	T	Intron
14223	-	T G	Intron
16915	-	G T	Beyond ORF(3')

Context:

DNA
Position

2584 ATAAACCGCTATCATTATTTATGCATCTAATCCTCTTGGGACCTGTTATCAGGTGAGCAA
CTTTTAAATCTTTTCCTTACCCCCCTAACCCCACCCCAGACTTGGGCAGAGAAAGATGAA
AGATTTACAAGATGGATACTATGGCTCTAATCAATTCTCTCATTTCCTCCCACTCTCGGC
TTCCCTGTCTACCATTCAGAAAACTTACCTGAAATCTTAAATGCCACCATGATGAACATG
TGGTATGTACTTGTGTTCCAAAACAATGAACGATGCTATTTGGGCTGTGTAAACTAGAAT
[G,C]
GGAACAACAAGACGTGATCACCCTGTGCATGAAGGCCATAGCTGCAGAGTGTGTAATTTT
ATTTAAAAAAATTTTTTTTTCTGAGACAAGGTCTTGCTCTGCCTCCCAGGCTACAGTGCA
GTGGTGCGATCATGGCTCACTGCAGCCTTGATCTCCTGGGATCAAGCGAACCTCCCACCT
CAGCCTCCAAGTAGCTGGGACCAAAGGAATGTGTCACCATGCCTGGTTAATTAAAAAAAA
ATTTTTATAGGCCGGGTGTGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAG
(SEQ ID NO: 20)

2655 TTTCCTTACCCCCCTAACCCCACCCCAGACTTGGGCAGAGAAAGATGAAAGATTTACAAG
ATGGATACTATGGCTCTAATCAATTCTCTCATTTCCTCCCACTCTCGGCTTCCCTGTCTA
CCATTCAGAAAACTTACCTGAAATCTTAAATGCCACCATGATGAACATGTGGTATGTACT
TGTGTTCCAAAACAATGAACGATGCTATTTGGGCTGTGTAAACTAGAATGGGAACAACAA
GACGTGATCACCCTGTGCATGAAGGCCATAGCTGCAGAGTGTGTAATTTTATTTAAAAAA
[A,T]
TTTTTTTTTCTGAGACAAGGTCTTGCTCTGCCTCCCAGGCTACAGTGCAGTGGTGCGATC
ATGGCTCACTGCAGCCTTGATCTCCTGGGATCAAGCGAACCTCCCACCTCAGCCTCCAAG
TAGCTGGGACCAAAGGAATGTGTCACCATGCCTGGTTAATTAAAAAAAAATTTTTATAGG
CCGGGTGTGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGTGGATC
ACCTGAGGTCAGGAGTTCAAGACCAGCTGGCCAACATGGTGAAACCCCTGTCTCTACTAA
(SEQ ID NO: 21)

FIGURE 3G

3693 TCTCTGTCGGTTTATCAGTTTCCTATTTATCTCTTTGTATATTTCTGCAATAAAGATACG
AAGTTGGGAGGGGGCAAAGGAAGGCAGTTCATCTCTCTATGTGGATGCAGTAGCACAATT
TAATAGTATCAAGTATTTCCATTCAGATTGCCTTGAAGTGGAAAGAATGCACTTAATCCT
AGCGAGATAGGCACCTGTGTCAACAGTCTCATCTGGATGCTATGGGGTTTTCAAGGTAGA
GAGATGTTGCAAAACTTATGAGTTCAGGAGTAAGGAATGGACCAAGTTTGTCTTGATTGC
[G,A]
AGAGAGGCAGACAACTGCAGTCAGCCGAGGAATATGGGTCAGAGTGTTGCAATGGGAAGA
TACCTCATCATTAGACAACTAAAAAGTCTGTGAAACTAATTAAGGATGGAACTCACTCCT
TTATAAAATTTCATATCTGTACACATGTATAATTTTTATTTGTCACTTATACCTCAATAA
GGCCAAAAAAAATTTTTTATCAATAAATTTTTAAGTGGGGAGGAATCGATTAGGCTCTATC
AGAGAGAATATGGGATATCAATGGAAACAGTGGCCTGAAATTTGGAGTCTAGTCTTCCGC
(SEQ ID NO: 22)

3992 CGAGAGAGGCAGACAACTGCAGTCAGCCGAGGAATATGGGTCAGAGTGTTGCAATGGGAA
GATACCTCATCATTAGACAACTAAAAAGTCTGTGAAACTAATTAAGGATGGAACTCACTC
CTTTATAAAATTTCATATCTGTACACATGTATAATTTTTATTTGTCACTTATACCTCAAT
AAGGCCAAAAAAAATTTTTTATCAATAAATTTTTAAGTGGGGAGGAATCGATTAGGCTCTA
TCAGAGAGAATATGGGATATCAATGGAAACAGTGGCCTGAAATTTGGAGTCTAGTCTTCC
[G,C]
CCTGTCATTGACTGGTTGTGTGTTCTTGGTAAAATCTCTGAAGATGGCTTCACAGGAAGG
CATATAGAGTTCCCTCATCTGTAAAGCAAATGGGTTAGTCTAAATCATGGGTCTCAAACT
CAAACACTTGCAGGGACCAGGCAGGTATCATAAATGAATGAAGCAGGCCTAGTATAAGAA
AAAACAGTAGCCTTGTGTGAGATGATAAATGGAAACAAAGTCTCAGAGAAATACTGAGGA
GTAGTGAGTACCATGGTAATCTGAAATCTTCATGACCTGCCTGAAGGAGGTAGCCCCTCT
(SEQ ID NO: 23)

6285 TCTTTCGAGGGCCTAGTTATGTAGTTCATTCAGGTTTGAGTTGTCGTCTTTTAAGTACTT
TTGTTGCTTTGATGGCTTCCTGTGTATATGAGATATTTTTTTTCCTCTGATCTGTCCCAA
GACTTTTTGGCTGAGATATGGTTGTGAGCCCTTTCTTGAAAAAGCAGAATCTGGCCAGGC
GCAGTGGCTCATGCCTGTAATCTCAGCACTTTGGGAAGCTGAGGTGGGTGGATCACCTGA
GGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAAACCCGTCTCTACTAAAAATAC
[-,A]
AAAAAAAAAAAAACCTTAGCCGGACATGGTGGCACATGCCTGTAATCCCAGCTACTCAGG
AGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCAGAGGTTACAGTGAGCTGAGATCG
CGCCAGTGCACTCCAGCCTGGGCGACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAAA
AGAAAGAAAGAAAAAGAAAAAGCAGAATCTAAAACTTTGGTTATGGAGCTGAATGCTTTG
AGGGAGGAATGCTTTACCTCACGAATTTGAGGTAAGAAAACAGGGCCTTTGGAACCTTCA
(SEQ ID NO: 24)

7066 TTGTAGAAGGCATGATCCACCCTTTGACTTATGAGAAATGATCAGAACAGAAGAGAGAAA
AAGACAAAAAGTAGTGCAGGCTGGCCATGGTGTCTCACACGTGTGATCCCAGCACTTTAG
GATCCCAGCACTTTGGGTCAAGGCAGTAGGATTGCTTGAGCCCAGGAGTTTGAGACCAGT
CTGGGCAACATGTCTAGATCTCCTCTCTACACAAATTAAAAATAGCTGGCATGGTGGCAT
GCGCCTGTAGTCCTAGCTACTCAGAAGGCTGAGGTGGGAGGATCATTTGAGCCTAGGAGG
[A,T]
CAAAGCTGCAATGAATTATGATTGTGCCACTGCACTCCAGCCAGGGTGATGGAGTAAGAC
CTTGTCTCAAAAATAAAATAAAGTAGCACAACCTCCCCAAGTTATTTTTTTCCCTCACTA
CAACCTCCCTTCCCAGGACAGCTTAGTTAAGTTTGCATGATGCTTTACTTCTGCAGATGT
TTGGAGGCCATGATTAAGTACCTCACACTGTGGAAGAAAGAGGAGCAGGAGGAGCCCTAT
GTCAGCCTCACCCGAAAGAAGATGCTAATAGATGGCGAGGAGGTGCTGATAGAATGGGAG
(SEQ ID NO: 25)

14223 AAGGAAGTTTATTTTGTATATTTATATGATTATTAAAGTGTTACAGTATATGTTCATCAT
GAGAAATTTAGAAAATAGAGAAATGTAGAGAAAAAGATTTCTAAAACTGATATAAGACTA
TCACACACAAAAAAAGATATTTTGGTTCATTTTTTCAATTTTTTGTGCATCTATTTTGTT
TTATTGTATATATTCAAGGTGTACAATGTGATGTTTCGATGTATGTACACATTGTGAAAT
GATTACCACAACCAAACTAATTAACACATTCATCACCTCACATAGTTATCATTTTTGTAC
[-,T,G]
TGTGTGTGTGTGTGTGTGTGTGTGTGTGGTAAAACTTAAGATCTACTCTCTTTAAAAATT
TCAAGTACACAATACATTATTGTCAACTATAGTCATCATGTTGTACATTAGAGCTCTGAA
ACTTATTTATCTTATAACTCTAAATTTGTAGCCTTTGATCAAAATCCTTCTATTTCCCTA

FIGURE 3H

AATCCCCATCCCCTGGTAACCACCCATTCTACTCTGTTGCTAGGTGTTCAACTTTTTTAG
ATTCCACATATAAGTAAGACAATGCAGTATTTTTCTTTATGTGTCTAGCTCATTTCACTT
(SEQ ID NO: 26)

16915 ATAGGCACACTCTAAGGAGAAAGTGCAGAGTAGAATTCCTTCAGGGCATAAGCCAAAATG
ACTCTTTTTCTCAGGGACCTGCATGGGCCTCCAGCTTGTCTATTGGAATTGTTAAGTGAA
GCCTCTCACTTAGTGCCTCATTAGCAGAGATTTCCTCCAACCCAGCTTTTCTGTGCTCTT
GGTATTTTACTACTTGATGTGGACCTCAGAGAAGCTGAACTGTAATTGAAAATGTTTCCG
ATGTGTGGAAGAAATGAAGACTGCTTTGTGTCTGCTGTTGTCCTGAGTATTTCATTAATG
[-,G,T]
GTGTGTGTGTGTGTGTGTGTGTGTGTGTATGTGTATGTGTGTAGGGAAGAAAGTAATA
ATGGCTGAGACATCACCTTCATGTTGTTTGCGATTGGGATGGGTGACTAACACTCCAAGG
TAGAGTGAAGGCAGAGGAGGGAAACAAGATCACATTAAATCATCATCAGTACTGGTTTCT
GCCTACAGGAGTTTACTTTTTTTTTTTTTCCTTTTTTGAGATGGAGTCTCGCTCTGTTTT
CTAGGCTGAAGTGCAGTGGTGTGATCTTGGCTCACTGCAGCCTCTGCCTCCTGGGTTCAA
(SEQ ID NO: 27)

FIGURE 3I